Comparative evaluation of temporal nodes Bayesian networks and networks of probabilistic events in discrete time[0]

S. F. Galán[1], G. Arroyo-Figueroa[2], F. J. Díez[1], and L. E. Sucar[3]

[1] Departamento de Inteligencia Artificial, UNED, Madrid, Spain
{seve, fjdiez}@dia.uned.es
[2] Instituto de Investigaciones Eléctricas, Cuernavaca, Mexico
garroyo@iie.org.mx
[3] ITESM - Campus Cuernavaca, Mexico
esucar@itesm.mx

Abstract. *Temporal Nodes Bayesian Networks* (TNBNs) and *Networks of Probabilistic Events in Discrete Time* (NPEDTs) are two different types of Bayesian networks (BNs) for temporal reasoning. Arroyo-Figueroa and Sucar applied TNBNs to an industrial domain: the diagnosis and prediction of the temporal faults that may occur in the steam generator of a fossil power plant. We have recently developed an NPEDT for the same domain. In this paper, we present a comparative evaluation of these two systems. The results show that, in this domain, NPEDTs perform better than TNBNs. The ultimate reason for that seems to be the finer time granularity used in the NPEDT with respect to that of the TNBN. Since families of nodes in a TNBN interact through the general model, only a small number of states can be defined for each node; this limitation is overcome in an NPEDT through the use of *temporal noisy gates*. [0]

1 Introduction

Bayesian networks (BNs) [7] have been successfully applied to the modeling of problems involving uncertain knowledge. A BN is an acyclic directed graph whose nodes represent random variables and whose links define probabilistic dependencies between variables. These relations are quantified by associating a conditional probability table (CPT) to each node. A CPT defines the probability of a node given each possible configuration of its parents. BNs specify dependence and independence relations in a natural way through the network topology. Diagnosis or prediction with BNs consists in fixing the values of the observed variables and computing the posterior probabilities of some of the unobserved variables.

Temporal Nodes Bayesian Networks (TNBNs) [1] and *Networks of Probabilistic Events in Discrete Time* (NPEDTs) [5] are two different types of BNs for

[0] Published in: *Proceedings of the 3rd Mexican International Conference on Artificial Intelligence (MICAI'2004)*, pages 498-507, Mexico City, Mexico, 2004. Springer, Berlin, Germany.

temporal reasoning, both of them adequate for the diagnosis and prediction of temporal faults occurring in dynamic processes. Nevertheless, the usual method of applying BNs to the modeling of temporal processes is based on the use of *Dynamic Bayesian Networks* (DBNs) [3, 6]. In a DBN, time is discretized and an instance of each random variable is created for each point in time. While in a DBN the value of a variable V_i represents the state of a real-world property at time t_i, in either a TNBN or an NPEDT each value of a variable represents the time at which a certain event may occur. Therefore, TNBNs and NPEDTs are more appropriate for temporal fault diagnosis, because only one variable is necessary for representing the occurrence of a fault and, consequently, the networks involved are much simpler than those obtained by using DBNs (see [5], Section 4). However, DBNs are more appropriate for monitoring tasks, since they explicitly represent the state of the system at each moment.

1.1 The Industrial Domain

Steam generators of fossil power plants are exposed to disturbances that may provoke faults. The propagation of these faults is a non-deterministic dynamic process whose modeling requires representing both uncertainty and time.

We are interested in studying the disturbances produced in the drum level control system of a fossil power plant. The drum provides steam to the superheater and water to the water wall of a steam generator. The drum is a tank with a steam valve at the top, a feedwater valve at the bottom, and a feedwater pump which provides water to the drum. There are four potential disturbances that may occur in the drum level control system: a power load increase (LI), a feedwater pump failure ($FWPF$), a feedwater valve failure ($FWVF$), and a spray water valve failure ($SWVF$). These disturbances may provoke the events shown in Figure 1. In this domain, we consider that an "event" occurs when a signal exceeds its specified limit of normal functioning.

Arroyo-Figueroa and Sucar applied TNBNs to the diagnosis and prediction of the temporal faults that may occur in the steam generator of a fossil power plant [2]. In this work, we describe the development of an NPEDT for the same domain. We also present a comparative evaluation of both networks.

This paper is organized as follows. Sections 2 and 3 give some details regarding the application to our industrial domain of TNBNs and NPEDTs, respectively. Section 4 explains the process of selection of the evaluation method for the domain and presents the results obtained from it for the two systems considered. Finally, Section 5 summarizes the main achievements of this work.

2 TNBN for this Industrial Domain

Arroyo-Figueroa and Sucar developed a formalism called *Temporal Nodes Bayesian Networks* (TNBNs) [1] that combines BNs and time. They applied this formalism to fault diagnosis and prediction for the steam generator of a fossil power

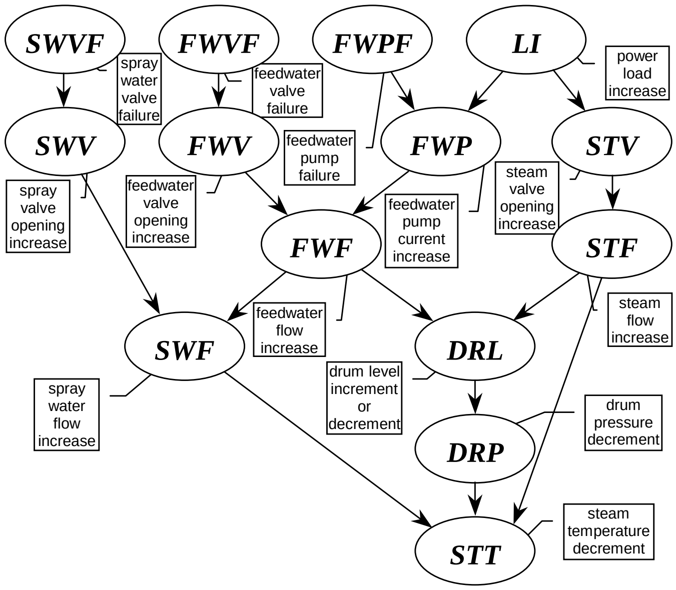

Fig. 1. Possible disturbances in a steam generator

plant [2]. A TNBN is an extension of a standard BN, in which each node represents a temporal event or change of state of a variable. There is at most one state change for each variable in the temporal range of interest. The value taken on by the variable represents the interval in which the change occurred. Time is discretized in a finite number of intervals, allowing a different number and duration of intervals for each node (multiple granularity). Each interval defined for a child node represents the possible delays between the occurrence of one of its parent events and the corresponding change of state of the child node. Therefore, this model makes use of **relative time** in the definition of the values associated to each temporal node with parents.

There is an asymmetry in the way evidence is introduced in the network: The occurrence of an event associated to a node without parents constitutes direct evidence, while evidence about a node with parents is analyzed by considering several scenarios. When an initial event is detected, its time of occurrence fixes the network temporally.

The causal model used by Arroyo-Figueroa and Sucar in the construction of their model is shown in Figure 1. The parameters of the TNBN were estimated from data generated by a simulator of a 350 MW fossil power plant. A total of more than nine hundred cases were simulated. Approximately 85% of the data were devoted to estimate the parameters, while the remaining 15% was used in the evaluation of the model, which we discuss later in the paper.

TNBNs use the general model of causal interaction, but lack a formalization of canonical models for temporal processes. Furthermore, each value defined for an effect node, which is associated to a determined time interval, means that the effect has been caused during that interval by only one of its parent events. However, this is not the general case in some domains where evidence about the occurrence of an event can be explained by several of its causes.

3 NPEDT for this Industrial Domain

In an NPEDT [5], each variable represents an event that can take place at most once. Time is discretized by adopting the appropriate temporal unit for each case (seconds, minutes, etc.); therefore, the temporal granularity depends on the particular problem. The value taken on by the variable indicates the **absolute time** at which the event occurs.

Formally speaking, a temporal random variable V in the network can take on a set of values $v[i]$, $i \in \{a, \ldots, b, never\}$, where a and b are instants —or intervals— defining the limits of the temporal range of interest for V. The links in the network represent temporal causal mechanisms between neighboring nodes. Therefore, each CPT represents the most probable delays between the parent events and the corresponding child event. For the case of general dynamic interaction in a family of nodes, giving the CPT involves assessing the probability of occurrence of the child node over time, given any temporal configuration of the parent events. In a family of n parents $X_1, \ldots, X_n$ and one child Y, the CPT is given by $P(y[t_Y] \mid x_1[t_1], \ldots, x_n[t_n])$ with $t_Y \in \{0, \ldots, n_Y, never\}$ and $t_i \in \{0, \ldots, n_i, never\}$. The joint probability is given by the product of all the CPTs in the network. Any marginal or conditional probability can be derived from the joint probability.

If we consider a family of nodes with n parents and divide the temporal range of interest into i instants, in the general case the CPT associated to the child node requires $O(i^{n+1})$ independent conditional probabilities. In real-world applications, it is difficult to find a human expert or a database that allows us to create such a table, due to the exponential growth of the set of required parameters with the number of parents. For this reason, temporal canonical models were developed as an extension of traditional canonical models. In this fault-diagnosis domain, we only need to consider the *temporal noisy OR-gate* [5].

3.1 Numerical Parameters of the NPEDT

In our model, we consider a time range of 12 minutes and divide this period into 20-second intervals. Therefore, there are 36 different intervals in which any event in Figure 1 may occur. Given a node E, its associated random variable can take on values $\{e[1], \ldots, e[36], e[never]\}$, where $e[i]$ means that event E takes place in interval i, and $e[never]$ means that E does not occur during the time range selected. For example, $SWF = swf[3]$ means "spray water flow increase occurred between seconds 41 and 60". As the values of any random variable in the network

are exclusive, its associated events can only occur once over time. This condition is satisfied in the domain, since the processes involved are irreversible. Without the intervention of a human operator, any disturbance could provoke a shutdown of the fossil power plant.

We use the temporal noisy OR-gate as the model of causal interaction in the network. In this model, each cause acts independently of the rest of the causes to produce the effect. Independence of causal interaction is satisfied in the domain, according to the experts' opinion.

Computing the CPT for a node Y in the network requires specifying

$$c_{y[k]}^{x_i[j_i]} \equiv P(y[k] \mid x_i[j_i], x_l[never], l \neq i)$$

for each possible delay, $k - j_i$, between cause X_i and Y, when the rest of the causes are absent. Therefore, given that X_i takes place during a certain 20-second interval, it is necessary to specify the probability of its effect Y taking place in the same interval —if the rest of the causes are absent—, the probability of Y taking place in the next interval, and so on. These parameters were estimated from the same dataset used by Arroyo-Figueroa and Sucar in the construction of the TNBN.

In the NPEDT, evidence propagation through exact algorithms takes, in general, a few seconds by using Elvira[1] and the factorization described in [4]. (If that factorization is not used, evidence propagation takes almost one minute.) Consequently, this network could be used in a fossil power plant to assist human operators in real-time fault diagnosis and prediction.

4 Evaluation of the TNBN and the NPEDT

A total of 127 simulation cases were generated for evaluation purposes by means of a simulator of a 350 MW fossil power plant [8]. Each case consists of a list

$$((event_1, t_1), (event_2, t_2), \ldots, (event_{14}, t_{14}))$$

where t_i is the occurrence time for $event_i$. There are 14 possible events, as Figure 1 shows. If $event_i$ did not occur then $t_i = never$. In general, among the 14 pairs included in each case, some of them correspond to evidence about the state of the steam generator.

4.1 Selection of the Evaluation Method

Our first attempt to quantify the performance of each model was carried out as follows. For each node or event X not included in the evidence:

1. Calculate $P^*(X) = P(X \mid \mathbf{e})$, the posterior probability of node X, given evidence $\mathbf{e}$.

[1] Elvira is a software package for the construction and evaluation of BNs and influence diagrams, which is publicly available at `http://www.ia.uned.es/~elvira`

2. For each simulated case that includes $\mathbf{e}$, obtain $ME(P^*(X), \hat{t}_X)$, a "measure of error" between the posterior probability and $\hat{t}_X$, the real (or simulated) occurrence time for X.
3. Calculate the mean and variance of the measures of error obtained in the previous step.

Given a *probability density function* for a variable V, $f_V(t)$, if V took place at $\hat{t}_V$, a possible measure of error is

$$ME(f_V(t), \hat{t}_V) = \int_0^{+\infty} f_V(t) \cdot \left| t - \hat{t}_V \right| \, dt. \tag{1}$$

This measure represents the average time distance between an event occurring at $\hat{t}_V$ and another one that follows distribution $f_V(t)$. For example, if $f_V(t)$ is a constant distribution between t_i and t_f (with $t_i < t_f$):

$$f_V(t) = \begin{cases} 0 & \text{if } t < t_i \\ \frac{p}{t_f - t_i} & \text{if } t_i \leq t \leq t_f \\ 0 & \text{if } t > t_f \end{cases} \tag{2}$$

then

$$ME(f_V(t), \hat{t}_V) = \begin{cases} p \cdot \left(\frac{t_i + t_f}{2} - \hat{t}_V \right) & \text{if } \hat{t}_V \leq t_i \\ \frac{p}{t_f - t_i} \cdot \left[\left(\hat{t}_V - \frac{t_i + t_f}{2} \right)^2 + \left(\frac{t_f - t_i}{2} \right)^2 \right] & \text{if } t_i \leq \hat{t}_V \leq t_f \\ p \cdot \left(\hat{t}_V - \frac{t_i + t_f}{2} \right) & \text{if } \hat{t}_V \geq t_f \end{cases} .$$

Note that, if time is the variable considered, ME is equivalent to the *prediction error* (difference between the observation at time t and the *expected forecast value* for time t). The probability distribution $f_V(t)$ can be directly obtained from $P^*(V)$ in an NPEDT, while in a TNBN it is necessary to know which parent node is really causing V, which can be deduced from the information contained in the corresponding simulated case.

Two problems arise when we try to apply Equation 1 to a node of either a TNBN or an NPEDT:

- If, given a simulated case, event V does not occur, we can only assign $\hat{t}_V$ the value $+\infty$; as a consequence, the integral in Equation 1 cannot be calculated.
- If $P^*(V = v[never]) > 0$, the value t in Equation 1 cannot be precisely defined for $V = v[never]$; if we supposed that $t = +\infty$, the integral could not be computed, as in the previous problem.

In order to avoid these two problems, we adopted an alternative point of view: Instead of a *measure of error*, a *measure of proximity* between $P^*(V)$ and $\hat{t}_V$ can be used for evaluating the networks. Given a probability density function for a variable V, $f_V(t)$, if V took place at $\hat{t}_V$, a possible measure of proximity is

$$MP(f_V(t), \hat{t}_V) = \int_0^{+\infty} \frac{f_V(t)}{1 + \left(\frac{t - \hat{t}_V}{c} \right)^2} \, dt \tag{3}$$

where c is an arbitrary constant. We have selected this function because it has four desirable properties:

1. As $\int_0^{+\infty} f_V(t)\ dt = 1$, $0 \le MP \le 1$. $MP = 1$ when $f_V(t)$ is a Dirac delta function at $\hat{t}_V$.
2. Note that when $t = \hat{t}_V$, the value of the integrand is $f_V(\hat{t}_V)$; however, as $\left|t - \hat{t}_V\right| \longrightarrow +\infty$, the integrand approaches 0 regardless of the value of f_V. The two following properties deal with the two problems presented above regarding the measure of error.
3. If, given a simulated case, event V does not occur ($\hat{t}_V = +\infty$), the integrand is zero when $t \neq never$ and we consider that $MP = P^*(V = v[never])$.
4. If $P^*(V = v[never]) > 0$ and $\hat{t}_V$ takes on a finite value, we consider that the contribution of $V = v[never]$ to MP is 0.
5. When the density function is constant inside an interval, MP can be easily calculated.

Since TNBNs and NPEDTs are discrete-time models, we calculate MP (given by Equation 3) by adding the contributions of each interval associated to the values of node V and the contribution of value *never*. $P^*(V)$ defines a constant probability distribution over each of the intervals defined for V. Given the constant distribution defined in Equation 2,

$$MP\left(f_V(t), \hat{t}_V\right) = \frac{p \cdot c}{t_f - t_i} \cdot \left(\arctan\left(\frac{t_f - \hat{t}_V}{c}\right) - \arctan\left(\frac{t_i - \hat{t}_V}{c}\right)\right).$$

As expected, the maximum measure of proximity appears when $\hat{t}_V = \frac{t_i + t_f}{2}$. We have used $c = 360$ in the TNBN and the NPEDT. In the NPEDT $t_f - t_i = 20$ seconds, while in the TNBN $t_f - t_i$ is specific of each interval.

4.2 Results

By using the measure of proximity proposed in Section 4.1, we have performed tests for prediction and for diagnosis from the 127 simulation cases available.

Prediction In order to analyze the predictive capabilities of the networks, we have carried out four different types of tests. In each of them there was only an initial fault event present: *SWVF*, *FWVF*, *FWPF* and *LI*, respectively. The states of the rest of the nodes in the networks were unknown. The time at which the corresponding initial fault event occurred defines the beginning of the global time range considered. Among the 127 simulated cases, 64 are associated to the presence of *LI*, and the rest of the initial fault events are simulated by means of 21 cases each. Tables 1 through 4 contain the means and variances of the measures of proximity obtained separately for both the TNBN and the NPEDT in the predictive tests. The average of the values shown in the last file of each table are: μ(TNBN) = 0.789003, σ^2(TNBN) = 2.603E-4, μ(NPEDT) = 0.945778, and σ^2(NPEDT) = 1.509E-3.

Table 1. Means and variances of *MP* when *SWVF* is present

Node	μ (TNBN)	σ^2 (TNBN)	μ (NPEDT)	σ^2 (NPEDT)
SWV	0.99786	2.185E-6	0.996066	3.456E-6
SWF	0.85793	4.267E-6	0.987375	4.138E-5
STT	0.55219	2.515E-3	0.874228	0.011258
Average	0.80266	8.404E-4	0.952556	3.767E-3

Table 2. Means and variances of *MP* when *FWVF* is present

Node	μ (TNBN)	σ^2 (TNBN)	μ (NPEDT)	σ^2 (NPEDT)
FWV	0.99844	2.053E-6	0.995963	1.286E-6
FWF	0.88154	8.99E-5	0.957003	2.39E-4
SWF	0.71559	1.069E-6	0.818828	4.704E-4
DRL	0.85165	1.035E-3	0.914457	0.003583
DRP	0.93127	6.317E-6	0.895225	0.006293
STT	0.14576	5.724E-8	0.832621	0.001401
Average	0.754041	1.89E-4	0.902349	1.998E-3

Table 3. Means and variances of *MP* when *FWPF* is present

Node	μ (TNBN)	σ^2 (TNBN)	μ (NPEDT)	σ^2 (NPEDT)
FWP	0.87001	3.522E-8	0.996962	7.425E-7
FWF	0.90496	1.478E-5	0.989113	2.018E-5
SWF	0.89533	1.537E-5	0.972282	1.598E-4
DRL	0.88665	5.776E-8	0.976043	9.658E-5
DRP	0.93262	9.382E-7	0.975946	6.43E-5
STT	0.14463	4.961E-8	0.954822	4.348E-4
Average	0.772366	5.205E-6	0.977528	1.294E-4

Table 4. Means and variances of *MP* when *LI* is present

Node	μ (TNBN)	σ^2 (TNBN)	μ (NPEDT)	σ^2 (NPEDT)
FWP	0.93043	9.377E-6	0.988255	2.594E-5
STV	0.99858	1.455E-6	0.997886	8.289E-7
FWF	0.83527	1.682E-5	0.97729	2.795E-4
STF	0.99694	8.829E-6	0.992117	1.35E-5
SWF	0.69901	2.722E-6	0.967232	6.314E-4
DRL	0.62306	5.715E-8	0.71745	2.147E-5
DRP	0.99204	1.433E-5	0.978515	1.271E-4
STT	0.540239	1.094E-7	0.986699	5.011E-5
Average	0.826946	6.712E-6	0.95068	1.437E-4

Table 5. Means and variances of MP when STT and one of its parents are present

Node	μ (TNBN)	σ^2 (TNBN)	μ (NPEDT)	σ^2 (NPEDT)
$SWVF$	0.701059	0.041	0.802444	0.118323
$FWVF$	0.449978	3.147E-3	0.742676	0.071108
$FWPF$	0.450636	3.064E-3	0.754429	0.062699
LI	0.655584	0.114252	0.995538	3.649E-5
SWV	0.762227	0.01292	0.801968	0.118188
FWV	0.446817	3.518E-3	0.742875	0.070954
FWP	0.428903	0.044449	0.688344	0.041785
STV	0.6529	0.115369	0.995781	1.091E-5
FWF	0.488658	0.04751	0.5607	0.075691
STF	0.630933	0.100964	0.999926	2.671E-9
SWF	0.609587	0.025558	0.999525	2.228E-7
DRL	0.562921	0.049913	0.510202	0.109071
DRP	0.809094	0.117564	0.662385	0.164828
Average	0.588448	0.052248	0.788984	0.064053

The results show that the NPEDT predicts more accurately than the TNBN. In general, the difference between the exactitude of predictions from the two networks grows as we go down in the graph. Both networks predict correctly that some events do not occur. Such events have been omitted in the tables.

Diagnosis The diagnostic capabilities of the TNBN and the NPEDT were studied on one type of test: The final fault event, STT (see Figure 1), was considered to be present. The occurrence time for STT established the end of the global time range analyzed. In this type of test, all the 127 simulated cases were used. Since in a TNBN the introduction of evidence for a node with parents requires knowing which of them is causing the appearance of the child event, in this type of test it was necessary to consider information from two nodes: STT and its causing parent. Table 5 includes the means and variances of the measures of proximity obtained in this test. Again, the NPEDT performs better than the TNBN.

Although in general the measures of proximity for diagnosis are lower than those for prediction, that does not mean that our BNs perform in diagnosis worse than in prediction. There is another reason that explains this result: If we had

- two different probability density functions, $f_V(t)$ and $f_W(t)$, the former sparser (more spread out) than the latter, and
- two infinite sets of cases, C_V and C_W, following distributions $f_V(t)$ and $f_W(t)$, respectively,

then, from Equation 3, MP would be lower on average for V than for W, since $t - \hat{t}_V$ is on average greater than $t - \hat{t}_W$. Therefore, although a BN yielded satisfactory inference results both for variable V and variable W, Equation 3 could in general produce different average MPs for V with respect to W. This

is taking place in our tests. For example, in the NPEDT we calculated the mean number of states per node whose posterior probability was greater than 0.001. While in the prediction tests this number was approximately 5, in diagnosis it rose to nearly 9. Anyhow, the measure of proximity defined in Equation 3 allows us to carry out a comparative evaluation of the TNBN and the NPEDT.

5 Conclusions

Arroyo-Figueroa and Sucar applied TNBNs to fault diagnosis and prediction for the steam generator of a fossil power plant. We have recently developed an NPEDT for the same domain and have carried out a comparative evaluation of the two networks. Our evaluation method is based on a proximity measure between the posterior probabilities obtained from the networks and each of the simulated cases available for evaluation. Since the faults that may occur in the steam generator of a fossil power plant constitute dynamic processes, the proximity measure takes into account time as an important variable. We have performed different tests in order to compare the predictive as well as the diagnostic capabilities of the TNBN and the NPEDT. The results show that in general the NPEDT yields better predictions and diagnoses than the TNBN. There are two main reasons for that: Firstly, the use of temporal noisy gates in the NPEDT allows for a finer granularity than in the case of the TNBN and, secondly, the definition of the intervals in a TNBN is not so systematic as in an NPEDT and depends strongly on the domain.

Acknowledgements

This research was supported by the Spanish CICYT, under grant TIC2001-2973-C05-04. The first author was supported by a grant from the Mexican *Secretaría de Relaciones Exteriores*.

References

1. G. Arroyo-Figueroa and L. E. Sucar. A temporal Bayesian network for diagnosis and prediction. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI'99)*, pages 13–20, Stockholm, Sweden, 1999. Morgan Kaufmann, San Francisco, CA.
2. G. Arroyo-Figueroa, L. E. Sucar, and A. Villavicencio. Probabilistic temporal reasoning and its application to fossil power plant operation. *Expert Systems with Applications*, 15:317–324, 1998.
3. T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5:142–150, 1989.
4. F. J. Díez and S. F. Galán. Efficient computation for the noisy MAX. *International Journal of Intelligent Systems*, 18:165–177, 2003.
5. S. F. Galán and F. J. Díez. Networks of probabilistic events in discrete time. *International Journal of Approximate Reasoning*, 30:181–202, 2002.

6. U. Kjærulff. A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence (UAI'92)*, pages 121–129, Stanford University, 1992. Morgan Kaufmann, San Francisco, CA.
7. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA, 1988.
8. A. Tavira and R. Berdón. Simulador de la central termoeléctrica Manzanillo II. Technical Report 19018, Instituto de Investigaciones Eléctricas, Mexico, 1992.